ABSTRACT

Methods for determining statistical models for predicting disease risks of a population are provided. Two types of data associated with members of the population are collected. The data may include both genetic and non-genetic types of data. A candidate statistical model is selected for calculating the disease risk. The model has a plurality of parameters and is a function of only one of the two types of data. A data weight is determined for each member of the population. Members having like data of the other type have like weights. The parameters of the model are optimized by fitting the collected data to the model taking into account of the weights.